Ligand & Structure-Based Descriptors

Schrödinger Suite 2006

Schrödinger Press

Revision A, April 2006

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Table 1.1.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Ligand and Structure-Based Descriptors

The Ligand & Structure-Based Descriptors panel provides a convenient interface to several Schrödinger programs, which are used to generate descriptors for a QSAR model. The focus is on generating descriptors for a set of ligands that are docked to a receptor. Three of the programs, Liaison, Prime MM-GBSA, and the eMBrAcE module of MacroModel, operate on the ligand and the receptor. The other two, QikProp and Ligparse, operate on the ligand only. The descriptors extracted from Liaison, Prime MM-GBSA, and eMBrAcE are energetic properties related to ligand binding. QikProp generates ADME properties, and Ligparse generates structure-based properties such as functional group counts. These descriptors can be used as input to the model-generation facility in Strike.

To run any of the descriptor generation tasks, you must have an installed and licensed version of the appropriate software: Liaison 4.0, Prime 1.5, MacroModel 9.1 (for eMBrAcE), and QikProp 2.5.

The results are collected as a set of descriptors in a comma-separated value (`.csv`) file. If you chose entries from the Project Table for input, the descriptors are imported into the project when the calculations finish. You can then select the relevant entries in the Project Table, and use Strike to build a QSAR model.

Note: The receptor and the ligands must be properly prepared beforehand. See the *LigPrep User Manual* for information on ligand preparation, and Chapter 4 of the *Glide User Manual* for information on protein preparation.

The Ligand & Structure-Based Descriptors Panel

The Ligand & Structure-Based Descriptors panel is used to generate ligand and structure-based descriptors for a structure-based QSAR model of ligand binding to a receptor. The panel is divided into three sections, which are described below.

To open the Ligand & Structure-Based Descriptors panel, choose Ligand & Structure-Based Descriptors from the Applications menu in the main window.

Setting Job Options

In the Job options section, you can set a job name and select a host as the defaults for all programs. You can also control whether jobs are run serially or simultaneously. The controls are described below.

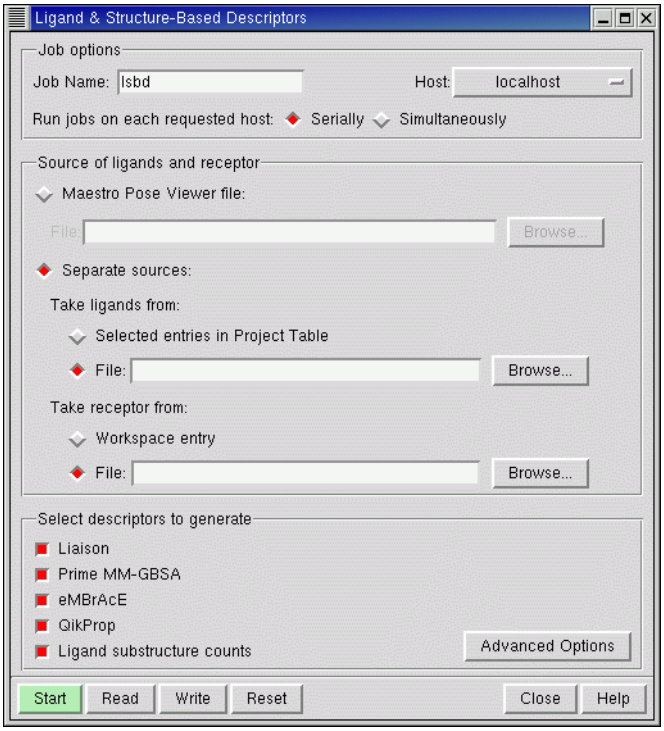

Figure 1. The Ligand & Strucutre-Based Descriptors ***panel.***

Job Name text box

Enter the name of the job in this text box. This name is used to name the input and output files for the job.

Host option menu

The Host option menu provides a list of hosts on which you can run the job. The choice you make from this option menu is the default host on which the computationally intensive jobs will be run. The master job and the ligand substructure counts job always run on the local host. For the jobs that generate the descriptors, this selection is superseded by any selection made in the Advanced Options panel. These jobs are run concurrently, and jobs for individual programs can be distributed over multiple processors if the program allows it. For the more time-consuming jobs, such as Liaison, eMBrAcE, and MM-GBSA, it is a good idea to choose different hosts for them in the Advanced Options panel.

Run jobs on each requested host options

These two options allow you to control whether jobs are run serially or simultaneously on each selected host. If you intend to run all the jobs on a single-procesor host, you should select Seri-

ally, so that the host is not overloaded. If you intend to run the jobs on a multiprocessor host, or submit them to a queueing system, or if you select different hosts for each descriptor types in the Advanced Options panel (see page 10), you should select Simultaneously to take advantage of concurrent execution.

Selecting the Ligands and the Receptor

To set up a calculation, you must select a receptor and a source of ligands. The receptor and ligands can come from a pose viewer file, or from separate files, or from the current project. The selections are made in the Source of Ligands and Receptor section.

Maestro Pose Viewer file options

If you want to use a pose viewer file for both the receptor and the ligands, select this option. To specify the file, enter the file name and path in the text box, or click Browse and navigate to the file. This option is useful if you have run Glide and want to obtain descriptors for a selected set of poses.

Separate sources options

If you choose to obtain the receptor and the ligands from separate sources, you can read them from files, or obtain them from the project.

To use the selected entries in the Project Table as the source of ligands, select Selected entries in Project Table, and select the ligands. When the job starts, the ligands are written to a Maestro file. To read the ligands from a file, select File and enter the file name and path in the text box, or click Browse and navigate to the file.

If the receptor is in the current project, you can display it in the Workspace and select Workspace entry. The receptor structure is written to a file when the job is launched. Otherwise you can read the receptor from a file. To do so, select File and enter the file name and path in the text box, or click Browse and navigate to a file.

Selecting the Descriptors To Generate

After choosing a receptor and ligands, you can choose the descriptors you want to generate. For each of the five descriptor sets, there is an option in the Select descriptors to generate section of the panel. You can select more than one option; the jobs will be run simultaneously.

The programs that generate the descriptors are run with default options that were chosen to produce reasonable descriptors. If you want to change any of these options or change some of the job options, click the Advanced Options button, and make your choices in the Ligand & Structure-Based Descriptors - Advanced Options dialog box. The settings available in this dialog box are a limited set of the full range of settings, designed to provide useful descriptors

for a QSAR model. For more information on the advanced options, see page 10. The descriptor types are described below.

Liaison descriptors

Liaison calculates ligand-receptor binding affinities using a linear interaction approximation runs molecular mechanics (MM) simulations of the ligand-receptor complex and of the free ligand and free receptor, with an SGB continuum solvation model. Three simulation methods are available: energy minimization (the default), molecular dynamics simulation, or hybrid Monte Carlo simulation. The latter explicitly include temperature effects. For more information on the Liaison model and methods, see Chapter 1 of the *Liaison User Manual*.

The Liaison descriptors include the eight components of the LiaisonScore, the LiaisonScore itself, and the five terms that contribute to the linear interaction model energy.

Prime MM-GBSA descriptors

The Prime MM-GBSA approach is used to predict the free energy of binding for a receptor and a set of ligands. MM-GBSA is an acronym for a method that combines OPLS molecular mechanics energies (E_{MM}), an SGB solvation model for polar solvation (G_{SGB}), and a nonpolar solvation term (G_{NP}) composed of the nonpolar solvent accessible surface area and van der Waals interactions. The total free energy of binding is then expressed as:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

where

$$G = E_{MM} + G_{SGB} + G_{NP}$$

The ligand in the unbound state is minimized in SGB solvent but is not otherwise sampled. In the calculation of the complex, the ligand is minimized in the context of the receptor. The protein is currently held fixed in all calculations. The following descriptors generated by the Prime MM-GBSA approach:

MM-GBSA_DG_bind	Ligand binding energy, ΔG_{bind}
MM-GBSA_E_complex	Energy of the complex, $G_{complex}$
MM-GBSA_E_protein	Energy of the receptor without the ligand, $G_{protein}$
MM-GBSA_E_ligand	Energy of the unbound ligand, G_{ligand}

eMBrAcE descriptors

eMBrAcE calculates ligand-receptor binding energies by molecular mechanics energy minimization of the complex and the separated receptor and ligand, with or without continuum solva-

tion. The eMBrAcE calculation is run in energy difference mode. The following descriptors are generated from the calculation:

Embrace_Total_Energy_without_constraints	Ligand binding energy
Embrace_Valence_Energy	Valence energy difference
Embrace_vdW_Energy	van der Waals energy difference
Embrace_Electrostatic_Energy	Coulomb energy difference
Embrace_Solvation_Energy	Solvation energy difference
Embrace_Constraint_Energy	Constraint energy difference

For more information on eMBrAcE, see Chapter 16 of the *MacroModel User Manual*.

QikProp descriptors

QikProp produces a list of 44 descriptors related to absorption, distribution, metabolism and excretion. These descriptors include properties like skin permeability and octanol/water partition coefficients, and counts of important functional groups. For a complete list of descriptors, see Chapter 1 of the *QikProp User Manual*

Ligand substructure counts

Counts of various substructures are generated from `ligparse`, and include counts of a wide range of functional groups (defined by single SMARTS pattern), counts of composite groups (defined by multiple SMARTS patterns), and some other counts. Over 100 functional groups are identified by `ligparse`. The following composite group counts are reported:

Num acceptor groups	Num acidic hydrogens
Num amide hydrogens	Num charged acceptor groups
Num charged donor groups	Num divalent oxygen atoms
Num donor groups	Num neutral acceptor groups
Num neutral amines	Num neutral donor groups
Num reactive groups	

and the remaining counts are of the following:

Num rings	Num heteroaromatic rings
Num aromatic rings	Num aliphatic rings
Num rotatable bonds	Num atoms
Molecular weight	Num chiral centers

Running the Jobs

When you have made all your selections, click Start to run the descriptor generation jobs. The input files are written, directories are set up for each kind of job, and the jobs are started. The calculations are run under Job Control, and can be monitored in the Monitor panel.

If you want to postpone running the jobs, save settings for use with other systems, or edit the input files to change options that are not available from the panel, you can click Write to write the input files without running the jobs. The main input file is written to *jobname*`.inp`, and is used to set up input files for the various programs. If you made job settings in the Advanced Options panel, these settings override those made in the main panel, and are stored in the input file.

To retrieve the settings from an input file, either to modify them or to run a job, click Read. A dialog box opens, in which you can navigate to and choose an input file, which has a file extension of `.inp`.

If you edit the the input files, you must then run the jobs for each program from the command line.

If you want to clear all custom settings and return to the default settings, click Reset. The calculation settings for all programs are returned to their defaults, which are as follows:

Liaison: Truncated Newton minimization with OPLS_2005 force field, 1000 steps, 15 Å residue-based cutoff, medium constraints option.

eMBrAcE: OPLS_2005 force field, no solvent, constant dielectric (1.0), normal cutoffs, PRCG minimization method with 5000 iterations, tiny constraints option.

QikProp: Normal mode.

Setting Advanced Options for Descriptor Generation

You can set nondefault options for the various programs that generate descriptors in the Ligand & Structure-Based Descriptors - Advanced Options panel. To open this panel, click Advanced Options. There are no options for ligand substructure counts.

Liaison Options

For Liaison, the options related to the calculation itself are the same as in the Parameters folder of the Liaison panel. For details, see Section 6.2.2 of the *Liaison User Manual*. These options provide most of the flexibility that is available for Liaison calculations.

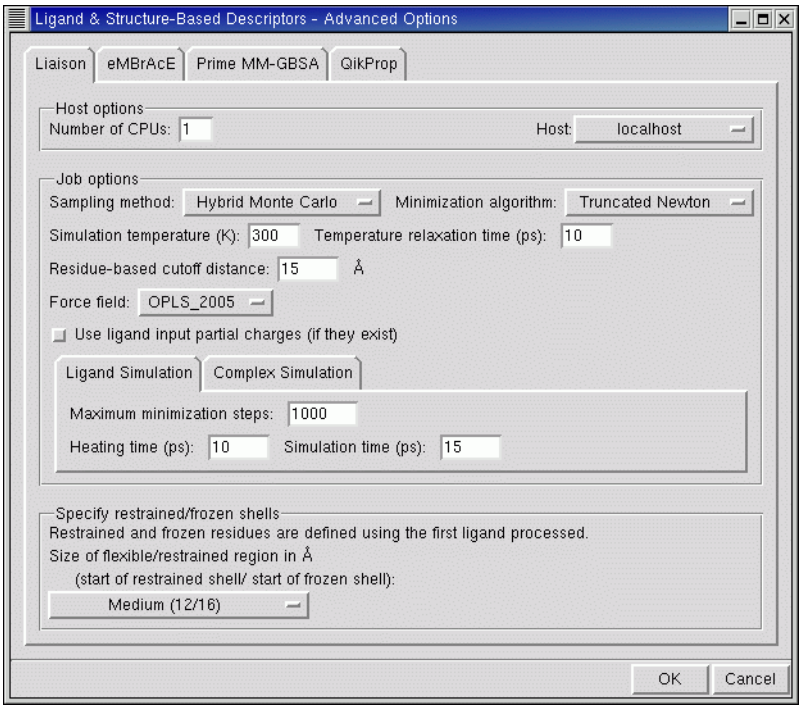

Figure 2. The Liaison ***folder of the*** Advanced Options ***panel.***

The Host options section provides a choice of hosts and number of processors over which to distribute the Liaison calculation.

If you choose a remote host or batch queue as the host for the job, ensure that the current working directory is mounted on the remote host. Liaison input files are written to the current working directory. Liaison does not have the ability to copy input files from a local directory to a remote scratch directory.

eMBrAcE Options

For eMBrAcE, the options for the potential and minimization are those from the Potential folder and Mini folder of the MacroModel panels. For more information on these options, see Section 5.2 and Section 7.2 of the *MacroModel User Manual*.

Constraints can be set by choosing from a menu that supplies a set of predetermined distances for restrained, frozen, and ignored shells. A Maestro file named *jobname*`-emb_cons.mae` is written and used to specify constraints relative to a single structure, so the calculations will be repeatable. From this file, the substructure`.sbc` file is written for each complex.

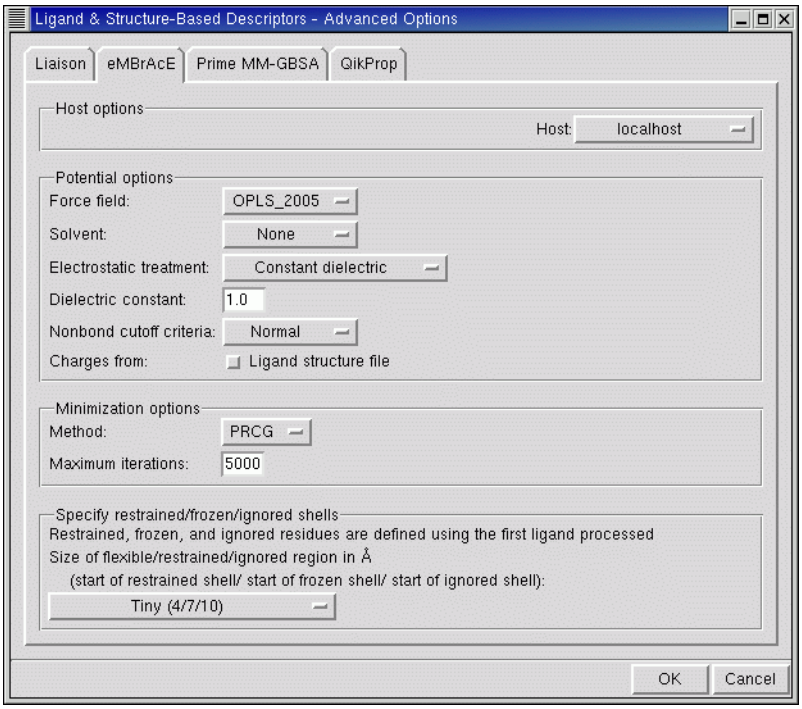

Figure 3. The eMBrAcE ***folder of the*** Advanced Options ***panel.***

Prime MM-GBSA Options

For Prime MM-GBSA, the only choices are the host options: the number of processors, the number of subjobs, and the host.

QikProp Options

For QikProp, the only program setting is whether or not to use fast mode. For a description of fast mode, see Chapter 1 of the *QikProp User Manual*. You can also select the host for the calculation.

Citing Ligand and Structure-Based Descriptors in Publications

Schrödinger Suite 2006 Ligand and Structure-Based Descriptors protocol; Liaison version 4.0, Schrödinger, LLC, New York, NY, 2005; MacroModel version 9.1, Schrödinger, LLC, New York, NY, 2005; Prime version 1.5, Schrödinger, LLC, New York, NY, 2005; QikProp version 2.5, Schrödinger, LLC, New York, NY, 2005.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the folder that is displayed in a panel, click the Help button in the panel. The Help panel is opened and a relevant help topic is displayed.
- For other information in the online help, open the Help panel and locate the topic by searching or by category. You can open the Help panel by choosing Help from the Help menu on the main menu bar or by pressing CTRL+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Liaison User Manual*, for information on Liaison
- *MacroModel User Manual*, for information on eMBrAcE
- *QikProp User Manual*, for information on QikProp
- *Strike User Manual*, for information on QSAR model building using Strike
- *Maestro User Manual*, for detailed information on using Maestro
- Frequently Asked Questions pages on the Schrödinger Support Center

The manuals are also available in PDF format from the the Schrödinger Support Center. Information on additions and corrections to the manuals is available from this web page.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- Liaison, Prime, MacroModel, or QikProp purchaser (company, research institution, or individual)
- Primary Liaison, Prime, MacroModel or QikProp user
- Computer platform type
- Operating system with version number
- Liaison version number
- Prime version number
- MacroModel version number
- QikProp version number
- Maestro version number
- mmshare version number

Copyright Notices

C and C++ Libraries for Parsing PDB Records

120 West 45th Street
32nd Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraße 13
68165 Mannheim
Germany

QuatroHouse, Frimley Road
Camberley GU16 7ER
United Kingdom

SCHRÖDINGER.